

01/26/01
1c971 U.S. PTO

01-29-01

Jc974 U.S. PTO
09/770541
01/26/01

The Assistant Commissioner of Patents
United States Patent and Trademark Office
Washington, D.C. 20231
ATTN: Box Patent Application

Re: U.S. Utility Patent Application
Appl. No. (Not yet assigned); Filed 1/26/01
For: **METHOD AND APPARATUS FOR AUTOMATICALLY GENERATING
HARDWARE FROM ALGORITHMS DESCRIBED IN MATLAB**
Inventor(s): **BANERJEE, et al**
Docket No.: **NWU-P001**

Sir:

The following documents are forwarded herewith for action by the U.S. Patent and Trademark Office:

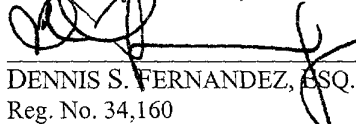
1. U.S. UTILITY APPLICATION
entitled: **METHOD AND APPARATUS FOR AUTOMATICALLY GENERATING
HARDWARE FROM ALGORITHMS DESCRIBED IN MATLAB**
having named inventor(s):
BANERJEE, et al
 - a. a specification consisting:
 - (i) 26 pages prior to the claims, including title page;
 - (ii) 4 pages of claims;
 - (iii) 1 page abstract;
 - b. 20 sheets of **informal** drawings: (FIGs. 1, 2a-c, 3, 4a-b, 5a-b, 6a-b, 7, 8a-b, 9, 10, 11a-c, 12-20);
2. An **original, executed** Combined Declaration and Power of Attorney by named inventors;
3. Form PTO-1082 (in duplicate);
4. Cover letter for Assignment (Form PTO-1595)
5. An original, executed Assignment to **NORTHWESTERN UNIVERSITY**, executed by named inventors, recordation of which is hereby requested;
6. A return post card; and
7. Check No. 1355 for \$ 1,710.00 to cover:

Patent application filing fee:	\$ 710.00
Assignment Recordation fee:	\$ 40.00
Excess claims fee:	\$ 960.00
8. Information Disclosure Statement and Form PTO1449

It is respectfully requested that the attached postcard be stamped with the filing date of the above documents and unofficial application number and returned to the addressee as soon as possible.

1-26-01
Date

Respectfully submitted,


DENNIS S. FERNANDEZ, ESQ.
Reg. No. 34,160

FERNANDEZ & ASSOCIATES, LLP
PATENT ATTORNEYS
PO BOX D
MENLO PARK, CA 94026-6204
(650) 325-4999
(650) 325-1203 : FAX
EMAIL: iploft@iploft.com

Certificate of Mailing By "U.S. Express Mail" Under 37 C.F.R. 1.10(c)

"EXPRESS MAIL" Mailing Label Number: EL759738719 US Date of Deposit: 1-26-01

I hereby certify that this paper and/or fee is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to the Assistant Commissioner For Patents, Washington, DC 20231

Name: Mantza Kidd 1-26-01

Signature Date Mantza Kidd Signature

Docket No.: **NWU-P001**

APPLICATION TRANSMITTAL LETTER

Assistant Commissioner of Patents
 United States Patent and Trademark Office
 Washington, D.C. 20231
ATTN: BOX PATENT APPLICATION

Sir:

Transmitted herewith for filing is the patent application of
 Inventor(s): **BANERJEE, et al.**
 Entitled: **"METHOD AND APPARATUS FOR AUTOMATICALLY GENERATING
 HARDWARE FROM ALGORITHMS DESCRIBED IN MATLAB"**
31 No. pages of specification, including title page, claims and abstract
20 No. sheets of X informal, _____ formal drawings

Also enclosed are:

- X Executed Combined Declaration and Power of Attorney for Patent Application
- X An Original Executed Assignment of the Application
- X Form PTO-1595 (Recordation Cover Sheet for Assignment)
- X An Information Disclosure Statement (Form PTO-1449A and Form PTO-1449B)
- X A copy of References cited in Information Disclosure: 3 documents

FEES DUE

 Applicant Claims Small Entity Status (37 CFR 1.27)

The fees due for filing the application pursuant to 37 C.F.R. 1.16 and for recording the Assignment, if any, are determined as follow:

CLAIMS					
	No. of Claims		Extra Claims	Rate	Fees
Basic Application Fee (\$710 large entity; \$355 small entity)					\$ 710.00
Total Claims	17	Minus 20 =		X \$18 =	.00
				X \$ 9 (small) =	
Total Independent Claims	15	Minus 3 =	12	X \$80 =	960.00
				X \$40 (small) =	
If Multiple Dependent Claims are presented, add \$260.00 or \$130.00(small)					
If Assignment enclosed, add Assignment Recording Fee \$40.00					40.00
TOTAL APPLICATION FEE DUE					\$ 1,710.00

PAYMENT OF FEES

The full fee due in connection with this communication is
 and is provided as follows:

\$ 1,710.00

_____ The Commissioner is hereby authorized to charge the fees associated with this communication or credit any overpayment to **Deposit Account No: 500482**. A duplicate copy of this authorization is enclosed.

X A Check No. 1355 for the above-specified full fee is enclosed. However, in case Applicant inadvertently miscalculated any required fee, the Commissioner is hereby authorized to charge the necessary additional amount associated with this communication or credit any overpayment to **Deposit Account No: 500482**. A duplicate copy of this authorization is enclosed.

NO 18 MONTHS PUBLICATION

REQUEST AND CERTIFICATION UNDER 35. U.S.C. 122(b) (2) (B) (i)

X I hereby certify that the invention disclosed in the attached application **has not and will not be** the subject of an application filed in another country, or under a multilateral agreement, that requires publication at eighteen month after filing. **I hereby request that the attached application not to be publish under 35 U.S.C. 122(b).**

This application is filed pursuant to 37 C.F.R. 1.53 in the name of the above-identified Inventor(s).

Please direct all correspondence concerning the above-identified application to the following address:

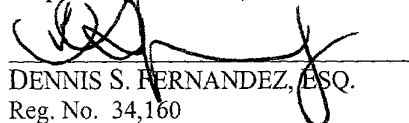
FERNANDEZ & ASSOCIATES, LLP
PATENT ATTORNEYS
PO BOX D
MENLO PARK, CA 94026-6204

(650) 325-4999
(650) 325-1203: FAX
EMAIL: iploft@iploft.com



22877
PATENT, TRADEMARK OFFICE

Respectfully submitted,


DENNIS S. FERNANDEZ, ESQ.
Reg. No. 34,160

1-26-2001
Date

109270-THS0260

5

Application

10

For

United States Non-Provisional Utility Patent

Title:

15

**METHOD AND APPARATUS
FOR AUTOMATICALLY GENERATING HARDWARE
FROM ALGORITHMS DESCRIBED IN MATLAB**

Inventors:

20

**Prithviraj Banerjee, residing at 2130 Chandler Lane, Glenview, IL 60025, a citizen
of United States of America.**

**Alok Choudhary, residing at 1100 N. Lake Shore Dr., #74, Chicago, IL 60611, a
citizen of the United States of America.**

25

**Malay Halder, residing at 2145 Sheridan Road, L458, Evanston, IL 60208, a citizen
of India.**

**Anshuman Nayak, residing at 2145 Sheridan Road, L458, Evanston, IL 60208, a
citizen of India.**

30

NWU-P001

METHOD AND APPARATUS FOR AUTOMATICALLY GENERATING HARDWARE FROM ALGORITHMS DESCRIBED IN MATLAB

5 STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with Government support by Defense Advanced Research
Projects Agency (DARPA) under Contract Number F30602-98-2-0144. The Government
may have certain rights in the invention.

10

Field of Invention

The invention relates to electronic design automation, particularly to the synthesis
of hardware from a high-level behavioral description.

15

Background of Invention

Certain high-level languages, such as MATLAB, are used for prototyping
algorithms in domains such as signal and image processing, simulation, analysis, etc. In
20 particular, MATLAB provides users with extensive libraries of high quality routines, as
well as high-level matrix-based syntax for expressing computations in a concise manner,
i.e., than available from conventional languages, e.g., C, Fortran.

However, because MATLAB is an interpretive language, programs thereof incur
25 high overhead during runtime. Thus, users developing applications for parallel

heterogeneous systems often prototype algorithms in MATLAB, then manually develop algorithms in C, assembly language for DSPs (Digital Signal Processors), embedded processors or in VHDL (VHSIC Hardware Description Language) or Verilog for synthesis and implementation on FPGAs (Field Programmable Gate Arrays) or ASICs (Application Specific Integrated Circuits). Such a manual process is tedious, inefficient, time-consuming, expensive, and unoptimal. Moreover, as hardware designs become faster and include more devices, improved software is needed for hardware synthesis.

Summary of Invention

The proposed novel electronic design tool and methodology enables automatic synthesis from programming languages, such as MATLAB. A MATLAB program is compiled into a high-level format, such as RTL-VHDL (Register Transfer Level - VLSI Hardware Description Language) or RTL Verilog, which is synthesized using computer-assisted tools to develop ASIC masks or FPGA configurations. Present methodology and system employs an array-oriented programming language such as MATLAB, having a large number of associated functions providing various constructs, such as operation on multi-dimensional arrays, function call statements, conditional statements, or loop statements.

Additionally, intermediate transformations and optimizations provide optimized RTL-VHDL and RTL Verilog description of given MATLAB program. Optimization may

include levelization, scalarization, pipelining, type-shape analysis, memory optimizations, precision analysis, or scheduling.

Brief Description of Drawings

5

Figure 1 shows flow chart of preferred method generally according to one aspect of present invention.

10

Figure 2 shows representative abstract syntax tree according to one aspect of present invention.

15

Figure 3 shows representative levelization according to one aspect of present invention.

Figure 4 shows representative translation of simple MATLAB program into finite state machine according to one aspect of present invention.

20

Figure 5 shows representative handling of conditional code for input MATLAB program according to one aspect of present invention.

Figure 6 shows representative handling of loops in input for MATLAB Program according to one aspect of present invention.

Figure 7 shows representative function call in MATLAB and function translation into corresponding state machine according to one aspect of present invention.

5 Figure 8 shows representative levelization and subsequent translation of array statement of MATLAB program according to one aspect of present invention.

10 Figure 9 shows representative finite state machine code section according to one aspect of present invention.

15 Figure 10 shows representative VHDL code generated for finite state machine according to one aspect of present invention. The corresponding Verilog code generation is similar.

20 Figure 11 shows representative real variables according to one aspect of present invention.

 Figure 12 shows representative loop unrolled for memory packing according to one aspect of present invention.

20

 Figure 13 shows representative overall framework for pipelining optimization according to one aspect of present invention.

Figure 14 shows representative terms for pipelining framework according to one aspect of present invention.

5 Figure 15 shows representative construction of nodes from MATLAB statements according to one aspect of present invention.

Figure 16 shows representative node construction for array access statements according to one aspect of present invention.

10 Figure 17 shows representative pipeline method according to one aspect of present invention.

Figure 18 shows representative construction of pipeline schedule from loop body
15 schedule according to one aspect of present invention.

Figure 19 shows representative renaming of scalars with live overlapping ranges
in
pipeline schedule according to one aspect of present invention.

20 Figure 20 shows representative VHDL code generation according to one aspect of present invention. The corresponding Verilog code generation is similar.

Detailed Description of Preferred Embodiment

Figure 1 shows overview of computer-automated electronic design or compilation process, which may be implemented in one or more software programs and computers or processing elements provided stand-alone or in network or otherwise distributed configuration. According to one operational mode of present automated approach, one or more digital circuit or system may be synthesized or otherwise defined from algorithm described in MATLAB or other programming language. Preferably, MATLAB program is compiled into high-level code, such as Register Transfer Level (RTL) – Very High Speed Integrated Circuit (VHSIC) High Definition Language (VHDL) or RTL Verilog, which is synthesizable using system-specific tools to develop Application Specific Integrated Circuit (ASIC) masks, Field Programmable Gate Array (FPGA) configurations, or other circuit implementations.

As described further herein, intermediate transformations and optimizations may be performed to obtain highly optimized description in RTL VHDL or RTL Verilog of a given MATLAB program. Additionally, optimizations include levelization, scalarization, pipelining, type-shape analysis, memory optimizations, precision analysis, scheduling, and other operations.

As shown in Figure 1, initially one or more input MATLAB program codes are parsed or otherwise processed using one or more directives to build one or more

Abstract Syntax Trees (AST) preferably according to the intermediate format (e.g.,
MATCH Intermediate Format (MIF)). Such an intermediate format may follow standard
syntax information based on MATLAB grammar, as well as one or more explicit or
implicit indications for design optimizations. Optionally, input code may contain user-
5 specified directives 36 regarding types, shapes, and/or precision of arrays. Directives 36
are attachable to MIF nodes as annotations.

Then, the type-shape analysis and inference phase 14 is applied. Because, by
default, MATLAB variables have no notion of type or shape, type-shape analysis phase
10 14 analyzes input program to infer type and shape of variables present. Next, the
scalarization phase 16 is applied, where operations on matrices may be expanded into
loops according to the internal format. When one or more optimized library functions is
available for a particular operation one of the library functions is used instead. Further,
after the scalarization step 16, levelization 18 may be applied, where one or more
15 complex statements are broken down into simpler representative statements. Scalarization
16 facilitates VHDL and Verilog code generation and/or optimizations.

Preferably, the transformation steps 12, 14, 16, 18 are performed on the MIF AST
format, and the output of such transformations is also in an MIF AST format. Moreover,
20 hardware-related optimizations may be performed subsequently on such MIF AST files.
For example, the precision analysis or inference scheme 20 is applicable to find the

minimum number of bits required to represent each variable in the MIF AST based on information available at compile time.

In addition, the memory optimization or transformation 22 may then be performed on MIF AST for optimization according to memory accesses present in the program as well as the characteristics of the external memory, i.e., when specified as an external input. Furthermore, the pipelining step 24 performs optimizations related to resources present and opportunities of parallel execution and pipelining available. Then, preferably, the MIF AST is translated 26 using RTL-VHDL or RTL Verilog grammar into an RTL-VHDL AST or an RTL Verilog AST 28. Finally, using one or more software intellectual property cores 32, tree traversal 30 of the optimized RTL-VHDL or RTL Verilog AST produces output code in RTL-VHDL or RTL Verilog 34.

The input MATLAB code is parsed using a formal grammar, and an abstract syntax tree is generated. Figure 2 shows a graphical view of (a) the hierarchy captured by representative grammar, (b) a sample code snippet, and (c) an abridged syntax tree for code snippet. Parsing and generation of AST are shown, such that MATLAB program is thereby provided using one or more .m" files. Each of such files may include one or more functions, wherein each listed function is defined per one or more statements listed. For example, each statement can contain "if ", "while" or "for" statement or a simple expression; each expression may correspond to an "atom", "operator", or a function call. Each atom can be constant or a variable. Figure 2 (b) shows a sample MATLAB

program, and Figure 2(c) shows the main components of the AST for the sample program.

Using user-specified directives 36, type-shape information can be provided to the present compilation process; and such directive information may be parsed for annotating the MIF AST. Hence, after the MIF AST is constructed, the compilation process invokes a series of phases, each phase processing the MIF AST, either by modifying or annotating the MIF AST with more information. Directives 36 serve as comments to the compiler, and thus may be used to allow user to provide more information to compiler about program to facilitate optimizations. For example, directive 36 may indicate when design information array includes items whose size at most will be a byte, such that compiler may optimize memory usage accordingly to reduced design space.

Using MATLAB-type program, type-shape analysis 14 of variables is accomplished effectively by carrying explicit data type and shape information, although MATLAB processing is generally interpretive, whereupon types of variables could be known at runtime before executing a statement. Hence, to compile and synthesize program written in MATLAB, such that maximum information about type and shape of arrays in particular, and of variables in general, are determined appropriately, algebraic framework is thereby provided to determine type and shape of arrays preferably at compile time. Representative directives (e.g., constraints, assertions, and hints) are provided as follow:

1. Constraint directives: State delay and throughput constraints at different levels of granularity. Constraint directives include resource constraint directives that specify resources available and their costs., e.g., `%!match DELAY 200ms` suggests to compiler maximum delay of 200 msec to complete an entire application task.
2. Assertions: Include assertions made about input MATLAB code, such as variable type assertions, value assertions, etc., e.g., `%!match BITS(32)` defines a 32-bit variable; helps invoke libraries for FPGAs or synthesize hardware with the right precision, but no more than necessary.
3. Hints: Suggestions to compiler, likely to improve performance, including parallelism hints, data distribution hints, platform preference hints, variable type/shape hints etc., e.g., `%!match DISTRIBUTE foo(CYCLIC(4), CYCLIC(4)) ONTO PROC(2,2)` defines distribution of the variable `foo` on a 2 x 2 processor mesh.

Scalarization 16 is applied to the MATCH intermediate format description for performing source-to-source transformation to a target language. In such step, the target language is typed statically, and only elemental operations are supported.

In preferred implementation, a high-level programming language is used, such as MATLAB, which is array-based, having built-in functions for supporting array operations. Moreover, to generate therefrom the low-level format, such as VHDL AST or Verilog AST, the corresponding MATLAB MIF AST is scalarized. Thus, to scalarize
5 MATLAB vector constructs, array shape and size are determined; although MATLAB is dynamically typed and may not ordinarily provide explicit basic data type and shape declarations. Accordingly, in accordance with one aspect of present invention, type-shape analysis 14 is applied.

10 Generally, translation is provided from one language having array constructs (e.g., MATLAB) to another language having loops and scalar operations (e.g., C), and scalarization may be performed upon intermediate format description (e.g., MIF-AST) to enable translation of array statements into loop form.

15 In particular, during operation of present methodology, given certain types and shapes of variables, for example, C-code may be generated to declare variables and corresponding statements. In this regard, compiler software may infer loop bounds for loops corresponding to vector statements provided preferably in MATLAB
Following is a sample MATLAB code:

20 $a = b + 2;$

where the correspondingly generated C code is:

float a[100; 200]; b[100; 200] ;

```

int i, j;

for(i = 0; i < 100; i++)

for(j = 0; j < 200; j++){

a[i][j] = b[i][j] + 2;

5      }

```

Preferably, the hardware description language, such as VHDL, is used for design file description for simulation and synthesis in accordance with present methodology; although certain constructs, e.g., file operations, assertion statements, or timing constructs may not be supported. Moreover, certain tools may require a specific coding style for generating hardware accurately. Hence, to enhance tool portability, the present methodology provides compiler that generates VHDL code that is compatible with various commercially-available high-level synthesis tools.

Furthermore, the VHDL AST format may be used, in addition to AST based on MATLAB grammar, to simplify final VHDL code generation, as well as enable hardware-related optimizations, like memory pipelining. Thus, during such optimizations, clock cycles and states may be introduced. Further, to generate VHDL AST, corresponding MATLAB AST is assumed to be scalarized, since MATLAB language is array-based.

Levelization 18 is applied to scalarized 16 MIF AST, modifying the AST to have statements in the three operand format only. Advantageously, different operators are spread across different states, so that optimal clock frequency is obtained, as shown, for example, in Figure 3. Levelization enables optimizations, such as operator chaining, resulting in a further optimized clock cycle. Since statements having large number of operations are broken down to a series of statements having one operation only, resources may be reused, as these smaller statements can be distributed across different clock cycles.

Scalarization and levelization steps 16, 18 transform input MATLAB code, so that such code includes a series of simple statements with constructs. Like conditionals, loops and function calls. Figures 4a-b show the transformation of a series of simple MATLAB statements into VHDL statements that are executed sequentially. The corresponding Verilog statements are similar but are not shown. Here, the state machine is synthesized by putting each simple statement in a state, and transitions between the states are arranged so that the states are traversed sequentially, i.e., one after another in order of their appearance in the MATLAB code. This sequencing results in modeling each state to operate in a clock cycle, while movement between the states is decided by the transition signal. Figure 10 shows the VHDL code generated for a representative finite state machine. The corresponding Verilog code is similar.

Next, during the synthesis flow, the compiler synthesizes one or more state machines traversing states for simple statements. For conditionals, a series of states is produced initially corresponding to the 'then' and 'else' body parts. A state is constructed to evaluate the condition, and transitions from the initial state are arranged so that states corresponding to the then-body are traversed when the condition is true, and states corresponding to else-body are traversed when the condition is false; see Figures 5a-b, for example.

Similar to conditional code, loops are handled such that the state machine is constructed for a body of the loop initially. Then, states are synthesized for initializing the index variable, incrementing such index variable, and checking exit condition of the loop. States are attached around the states for loop body, as shown in Figures 6a-b.

Moreover, in the synthesis process, each function call in the MIF AST is mapped to a state machine in the VHDL or Verilog AST; Figure 7 shows the state machine representation for a MATLAB code with a function call. Each function is declared as a process, and the arguments of a function are declared as signals. The function arguments are passed by assigning variables at the calling site to signals corresponding to arguments of the function. To assign variables at the caller site, signal names corresponding to arguments of the function and their ordering are known a priori. In this manner, an earlier pass is made generating the symbol table entry corresponding to each function definition, assigning unique names to signals corresponding to the arguments. Each function has an

in:start signal and an out:done signal. The execution of the function is started by calling the function by assigning values to the signals corresponding to the arguments and making high the in:start signal for the called function. The calling function waits for the out:done signal of called function to be high, after which the output signal of the called function holds valid values. Hence, advantageously, resources are shared between each function call, and the present approach is applicable to exploit functional parallelism. Preferably, since different processes may run concurrently, multiple processes may not write to shared signals simultaneously.

Present compiler declares scalars as variables to facilitate movement of operations across states by optimization phases. Variables corresponding to function arguments are declared signals to be visible outside the process corresponding to the function. Other signal declarations include signals corresponding to memory interface.

Furthermore, the compiler may map arrays to memory; specification of memory access characteristics is provided as an input. The compiler instantiates registers for scalars, e.g., on FPGAs. The levelization phase ensures that each statement has at most one memory access with no other associated operations. The exact mechanism and signals involved in accessing memory is specified in a file read by the compiler, which uses such information to produce states to read/write memory corresponding to each array access that appears in levelized and scalarized MATLAB code; Figures 8a-b shows an example.

Precision analysis 20 determines the minimum number of bits required to represent a variable. Since number of required bits relates to maximum and minimum value that variable can attain through program run, precision analysis 20 can be performed by value range propagation. Levelization serves to formulate series of transformations applicable on statements to infer the value ranges.

Moreover, real variables are represented in a way such that operations are accomplished using integer operators; both operands for any operator are integer or real. In particular, to avoid converting induction variables inside loops to be type promoted to real numbers, so-called temporaries are used. Because the MATLAB language is typed dynamically, without ordinarily representing type and shape of variables, data flow graph is used with single assignment property.

Precision analysis 20 uses an array-based single static assignment (SSA) representation where each array element that is written into more than once is renamed. Advantageously, increase in the value range of an individual array element does not increase the value range of the entire array, so that precision inferencing becomes more accurate. Precision analysis phase 20 ends once value range of all the variables stabilize. Precision information can be derived from target architecture for which VHDL is generated. Value range propagation benefits optimization approaches, such as constant propagation and dead code elimination.

Preferably, on reconfigurable computing platforms, fixed point representations may be used, since the dynamic range of variables in image and signal processing applications is relatively small. Further, real number representations are scaled down to a value between -1 and +1 so that the number of bits required to represent a real number is related directly to its resolution or number of digits after decimal point.

Figure 11a shows a MATLAB code 62 for multiplication of two real numbers.

Figure 11b shows the normal representation code 64 if both numbers are scaled down by the largest integer value of 255 to get the value within -1 and +1; the number of decimal bits needed to represent the transformed number may be as high as 32 bits, i.e., to limit error in calculating the result, resulting in instantiation of a 64-bit integer multiplier.

Further, since variables in the input code have to be scaled down by the maximum integer, this approach results in real variables requiring 32 bits leading to a large consumption of processing resources. Thus, in accordance with one aspect of the present invention, real numbers are represented by integer and fractional parts. Figure 11c shows the resulting transformed code. Transformation results in instantiation of a 13-bit multiplier, with no error in output calculation.

As described herein, the number of bits required to represent the integral part of a real number can be deduced from the precision analysis algorithm based on value range propagation. Resolution or minimum number of bits required for the fractional part can

be inferred after the error analysis phase. Preferably, real variables have the same number of bits for the fractional part; the number of resolution bits for real numbers is inferred when user specifies using directives; user uses output statement, (e.g., printf,) and defines output resolution; or compiler assumes that since the code was to be executed as sequential MATLAB code which has a default resolution of 4, output variables have a resolution of 4, and back propagate such information in error analysis phase to determine resolution of intermediate real variables. Foregoing analysis provides minimum number of bits required to represent fractional part of real numbers, while precision analysis algorithm in previous section provides minimum number of bits required to represent integer part of real number.

Additionally, optimal packing order (PO) algorithm is provided for each array, where PO is defined by the maximum number of array elements that can be packed in each memory location. The minimum number of bits required by array elements can be inferred from precision analysis. Since most of images read from MATLAB are stored in 2-dimensional arrays, the precision of input images is inferred by parsing input matrices to obtain the maximum value of various array elements. Figure 12 shows a loop described in MATLAB. Since memory packing involves unrolling the loop to find more consecutive array element accesses, dependence-analysis phase may be used to determine any loop carried dependencies.

Preferably, for memory optimization 22, memory packing is performed on the innermost loop of a deeply nested loop or innermost dimension of array access, and thus, analysis can be done by the greatest common denominator test (GCD). Since memory packing requires consecutive array accesses across loops, array access patterns are

5 determined across loop iterations. Unroll factor, i.e., number of statements unrolled, of each memory access in a loop is defined by the number of array element accesses across loops located in the same physical memory location. To minimize number of memory accesses, the loop is unrolled by the maximum unroll factor.

10 Additionally, pipelining 24 optimizes the number of cycles taken by a design to execute input application, as shown in Figure 13. Upon input of a MATLAB loop statement 70, the given series of nested loops, check 72 is performed on innermost loop body to determine if the pipelining method is applicable. If it is determined that the inner loop body is suitable for pipelining, then the pipelining algorithm is applied 74. Initially,

15 the inner loop body is located in the AST, then the nodes are constructed corresponding to statements in the loop body. Predicated nodes are constructed for conditional statements present in the loop body. A data flow graph utilizing nodes corresponding to statements of the loop body is constructed 76. Scheduling algorithm is applied 78 to the data flow graph. The schedule for loop body is used 80 to construct a schedule for the

20 pipeline; scalars with overlapping live ranges in the pipeline schedule are renamed. Loop conditionals are produced 82 and VHDL or Verilog statements are generated 84 from the pipeline schedule.

Generally, the pipelining 24 step attempts to pipeline innermost loop in sequence of nested loops, according to two conditions: loop under consideration is innermost loop; and no statement in the loop body depends on data defined by a statement in an earlier iteration, but appears after inner loop body. Body of loop statement includes other statements, which may be of three: simple assignment statements, conditional statements, and loop statements, as shown in Figure 14. Traversal of the AST is performed, and each loop statement is checked for nested loops. Simple assignment statements in the loop body are ignored. If a statement in the loop body is a conditional statement, then the body of the conditional statement is recursively traversed to check for the presence of loops.

If a loop statement is found in the loop body or by recursively traversing conditional statements in the loop body, the loop is judged to be an outer loop, and pipelining is not applied to such loops; else if no loop statement is found in the loop statement body or by recursively traversing conditional statements present in the loop body, then the loop is considered to be an innermost loop. Loops that originate from scalarization of matrix operations are marked to indicate that they do not have dependencies where statement in loop body depends on data defined by statements in earlier iterations, but appears after in loop body. For loops that do not originate from scalarization of matrix operations, GCD test is performed to check for the presence of dependencies.

Statements of the loop body are traversed one by one, and a node is constructed corresponding to each statement. Nodes are connected by dependency edges to form a dataflow graph. If conditional statements are present in a loop body, then a check is performed on the body of the conditional statement to ensure statements inside the conditional statement body do not modify any conditional variable of the conditional statement. If statements inside the body of a conditional statement modifies any conditional variables, then pipelining 24 is terminated. For statements inside the body of a conditional statement, nodes are created with predicates, e.g., 15.

During VHDL code generation corresponding to a particular node, produced VHDL code is guarded effectively by predicate expressions of such a node. For nodes corresponding to statements in the *true* path of the conditional statement, the predicate expression is the condition variable. For nodes corresponding to statements in the *false* path of the *false* conditional statement, the predicate expression is the negation of the condition variable. In case of nested conditional statements, the predicate expressions from higher nesting are concatenated to form the predicate expression of the node. For statements with array accesses, the procedure is slightly different; for array access statements, location of variable is computed first, i.e., for address calculation.

Then, after address is calculated, the series of signals are assigned specific to the memory interface in use. Given a multi-dimensional array access, a node is generated corresponding to the address calculation in each dimension. Signals assigned for memory

access are specified in an external file read by compiler, and nodes are generated corresponding to each state defined in the external file.

Furthermore, to construct the dataflow graph, an auxiliary control flow graph is constructed initially. In the control flow graph, node “x” is made a predecessor of another node “y”, if an execution path exists starting from the first node of the control flow graph that reaches the node “y” with node “x” immediately before in the path. After the control flow graph is constructed, for each node variable that the node defines, and the variables that the node uses are thereby determined. For each variable used by the node, the control flow graph is traversed upward, and all reaching definitions are located. A dependency edge is added from the node using the variable to all nodes with reaching definitions; such operation is applied to all nodes, and nodes along with the dependency edges define the dataflow graph.

The scheduling process is applied to the data flow graph, and assigns each node in the data flow graph a state number, then the initiation interval for the pipeline is determined. Initiation interval for a pipeline is the number of clock cycles between the initiation of consecutive iterations. Nodes correspond to statements of a loop body with state number assignments, is referred to as the schedule of the loop body. Nodes not dependent on any other nodes are considered initially for scheduling, and assigned state 0. For a given node, once all the nodes that the node is dependent on are scheduled, such node is ready to be scheduled, and such node is assigned the current state number.

When all the nodes that are ready in a step are assigned, then the state number is incremented to the next value. Exception occurs while assigning a state to a node corresponding to a memory access. If the node corresponding to a memory access is ready, such node is not assigned immediately the current state number. For nodes corresponding to memory accesses, the state number is determined such that it is closest to the current state number, and that the state number modulo the number of memory accesses in loop body is different from all state numbers modulo the number of memory accesses in the loop body corresponding to memory access nodes for which states have been assigned at that point. Initiation rate of pipeline is set to number of memory accesses in loop body. An example of the process in work is shown in Figures 14-17 showing representative construction method statements.

In Figure 17, the sample dependence graph of loop body to be pipelined is shown at left. Dark vertices denote memory references, while light nodes denote non-memory reference vertices; the initiation rate is 2. After placing first memory reference in state 0, second memory reference cannot be placed in state 4, although predecessors are assigned; This constraint is because $4 \bmod 2$ is 0, and $0 \bmod 2$ is also 0, which is assigned. So, the second memory reference is pushed to 5.

20

After the scheduling process assigns state numbers to all the nodes, pipeline is constructed. Here, L/I copies of the loop bodies are created, where L is length of the loop

body schedule, and I is the initiation interval of the pipeline; see Figure 18 for the representative pipeline schedule. L is defined by the largest state number assigned to any node, and I is equal to the number of memory accesses in the loop body; index variable corresponding to the loop is var . In i th copy of the loop body, var is replaced by $(\text{var} + i)$, then copies of the loop body are concatenated with an interval of I between the successive copies.

Next, all scalar variables in pipeline schedule are located, and the nodes defining scalars and the nodes using scalars are determined. States between the definition and use of scalars constitute scalar live range. Live range of each variable in each copy of the loop body that comprise the pipeline schedule is determined. Scalars are located for which the live range in one copy overlaps with the live range in another copy of the loop body. A new version is then created for such scalars for each copy of the loop body. Statements that define or use scalars with overlapping live ranges are converted into case statements. For i th case, $(i + j)$ th instance of the scalar variable is used in j th copy of the loop body, for example, as shown in Figure 1. A variable is defined that acts as counter starting with 0 till $\text{ceil}(L/I - 1)$.

Moreover, states from 0 to $L-I-1$ of the pipeline schedule comprises the prologue of the pipeline; states from $L - I$ to $L - 1$ comprise the kernel of the pipeline. The rest of the states are the epilogue of pipeline. Index variable and modulo variable are initialized at beginning of the pipeline kernel. Modulo variable is incremented at the last state of

kernel. The index variable is incremented till $n - \text{Ceil}(L/i) + 1$, where n is bound of the origin at loop. If index variable is less than $n - \text{Ceil}(L/i) + 1$, the state machine loops back to the first statement of the kernel; else the state machine jumps to the first statement of the epilogue.

5

Once the pipeline schedule is constructed, VHDL or Verilog code is generated from the schedule and added to the VHDL or Verilog AST. For each node, the basic statement is VHDL. Predicate list of the node is checked, and if predicate expressions exist, then the expressions are ANDed to form a single condition, which guards the execution of the basic statement of the node. All nodes assigned a state are associated in a single state of VHDL AST; see Figure 20, for example. The last statement of the kernel has a conditional statement depending on index variable count that decides the next state. For the rest of the states, the next state is the state that follows immediately.

15 Foregoing described embodiments of the invention are provided as illustrations and descriptions. They are not intended to limit the invention to precise form described. In particular, it is contemplated that functional implementation of invention described herein may be implemented equivalently in hardware, software, firmware, and/or other available functional components or building blocks. Other variations and embodiments are possible in light of above teachings, and it is thus intended that the scope of invention not be limited by this Detailed Description, but rather by Claims following.

Claims

What is claimed is:

1. Computer-automated electronic design methodology comprising the steps of:
accessing a first file comprising an algorithm that is described in an array-oriented
5 programming language such as MATLAB; and generating a second file comprising a
digital circuit representation that is synthesized automatically from the algorithm, where
the representation is Register Transfer level VHDL or Verilog.

2. The methodology of Claim 1 wherein:
10 the second file comprises a design feature for optimization that is described in an
intermediate format.

3. The methodology of Claim 1 wherein:
the second file is generated using a user directive, effectively allowing a user to
15 guide a compiler by specifying a variable type or shape for optimization.

4. In an electronic design system comprising a compiler for synthesizing a circuit
definition from a high-level definition, a process comprising the step of:
determining by a compiler a type or a shape of a variable or an intermediate
20 temporary variable.

5. Electronic design scalarization process comprising the step of:

transforming a first intermediate-format description into a second intermediate-format description by translating an array statement in the first intermediate-format description into a loop in the second intermediate-format description.

5 6. In an automated design system comprising a processor for optimizing synthesis, a precision-inferencing method comprising the step of:

determining by a processor a maximum bit precision for a variable for effecting a resource optimization.

10 7. In an electronic design automation system comprising a compiler for processing a file including a real variable, an error-analysis method comprising the step of:

determining a minimum number of bits for representing a real variable.

15 8. Compact electronic storage technique comprising the steps of:

identifying a plurality of arrays having a bit precision that is less than an available memory width; and

packing more than one array element into a memory location associated with the available memory.

20 9. In an electronic design automation system, a levelization method comprising the step of:

transforming a first statement in an intermediate-format description
into a second statement associated with only one operation.

10. In an electronic design automation system, a memory-access method
5 comprising the step of:
determining a memory access pattern or a memory access constraint for a design
synthesis.

11. In an electronic design automation system, a transformation method
10 comprising the step of:
transforming an intermediate format description for synthesizing an optimized
memory access.

12. In an electronic design automation system, a method comprising the step of:
15 determining a number of states in a finite state machine realization associated with
a transformed intermediate format description.

13. In an electronic design automation system, a method comprising the step of:
determining a finite state machine realization of one or more levelized statement
20 in an intermediate format description.

14. In an electronic design automation system, a method comprising the step of:

determining a high-level variable or signal assignment for one or more variable in
an intermediate format description.

5 15. In an electronic design automation system, a method comprising the step of:
identifying a pipeline opportunity in an intermediate format description.

10 16. In an electronic design automation system, a method comprising the step of:
transforming an intermediate format description to enable pipelining in a
produced high-level code.

15 17. In an electronic design automation system, a method comprising the step of:
producing a pipelined high-level code from a transformed intermediate format
description.

METHOD AND APPARATUS FOR AUTOMATICALLY GENERATING HARDWARE FROM ALGORITHMS DESCRIBED IN MATLAB

Abstract

5

Digital circuit is synthesized from algorithm described in the MATLAB programming language. A MATLAB program is compiled into RTL-VHDL, which is synthesizable using system-specific tools to develop ASIC or FPGA configuration.

10

Intermediate transformations and optimizations are performed to obtain highly optimized description in RTL-VHDL or RTL Verilog of given MATLAB program. Optimizations include levelization, scalarization, pipelining, type-shape analysis, memory optimizations, precision analysis and scheduling.

15

20

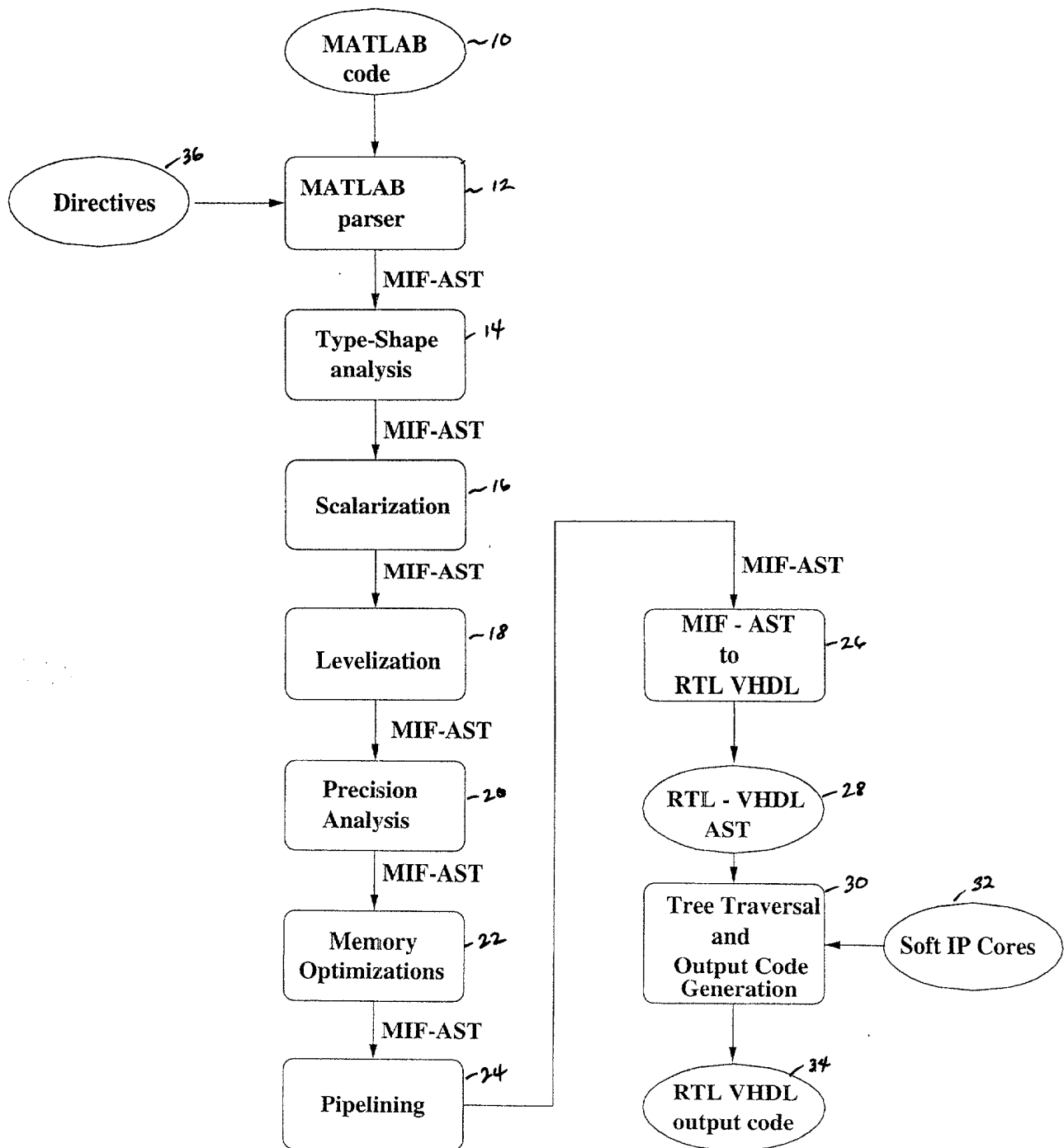
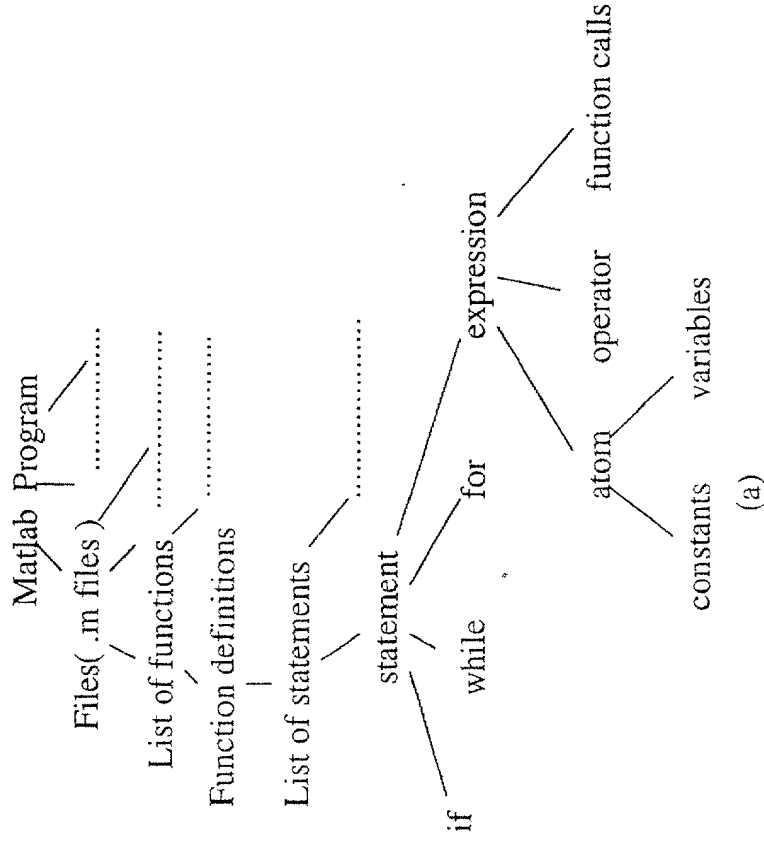


Figure 1: Synthesis flow.



```

pc_filter = rand(63,1);
pc_filter_time = zeros(9512,1);
pc_filter_freq = fft(pc_filter_time);
for i = 1:6
    pc2(i,:) = pc1(i,:)*pc_filter_freq
end;
  
```

(b)

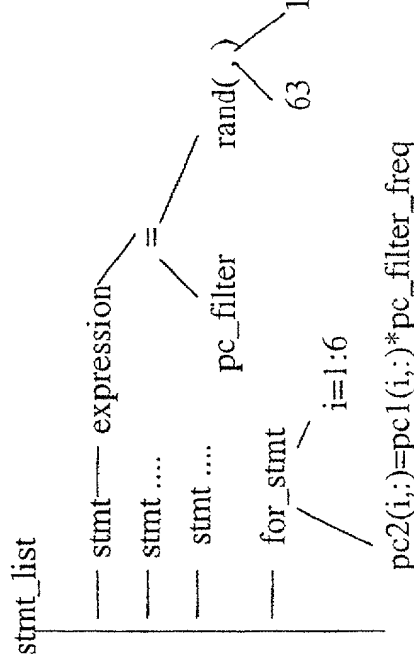


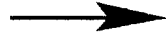
Figure 2: Abstract Syntax Tree: (a) The hierarchy captured by the formal grammar (b) A sample code snippet (c) Abridged syntax tree for the code snippet.

Unlevelized

a = b * c * d;

j = j + a;

i = i + 1;



state 1 : a = b * c * d;

state 2 : j = j + a ;

state 3 : i = i + 1 ;



clock period determined
by the longest state

Execution time = 3 X 30 = 90ns

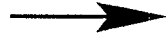
Levelized

t1 = b * c ;

t2 = t1 * d;

j = j + t2 ;

i = i + 1 ;

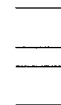


state 1 : t1 <= b * c ;

state 2 : t2 <= t1 * d ;

state 3 : j <= j + t2 ;

state 4 : i <= i + 1 ;



longest state reduced
by levelization

Execution time = 4 X 15 = 60ns

Time to multiply : 15ns

Time to add : 10ns

Execution time = Number of States X Clock Period

Figure 3: Levelization improves clock period.

```

a = 1;
b = 1;
c = a + b;
c = c + 1;

when state 1 => a <= 1;
               next_state <= state 2;

when state 2 => b <= 1;
               next_state <= state 3;

when state 3 => c <= a + b;
               next_state <= state 4;

when state 4 => c <= c + 1;

```

(a)

(b)

Figure 4: (a) Simple MATLAB statements. (b) State machine that steps through the statements sequentially.

```

if( x )
    a = b + 1;
else
    a = b - 2;
end;
c = a + 1;

when state 1 => if( x ) then
    next_state <= state 2;
else
    next_state <= state 3;
when state 2 => a <= b + 1;
    next_state <= state 4;
when state 3 => a <= b - 2;
    next_state <= state 4;
when state 4 =>
    c <= a + 1;

```

(a) (b)

Figure 5: (a) Conditional MATLAB code (b) State machine for Conditional Code.

```

for i = 1 : 256
    a = a + i;
end;

when state 1 => i <= 1;
    next_state <= state 2;

when state 2 => if( i <= 256 ) then
    next_state <= state 3;
else
    next_state <= state 5;
endif;

when state 3 => a <= a + i;
    next_state <= state 4;

when state 4 => i <= i + 1;
    next_state <= state 2;

when state 5 => [ next statement after loop ]

```

(a) (b)

Figure 6: (a) A MATLAB *for* loop (b) State machine for *for* loop.

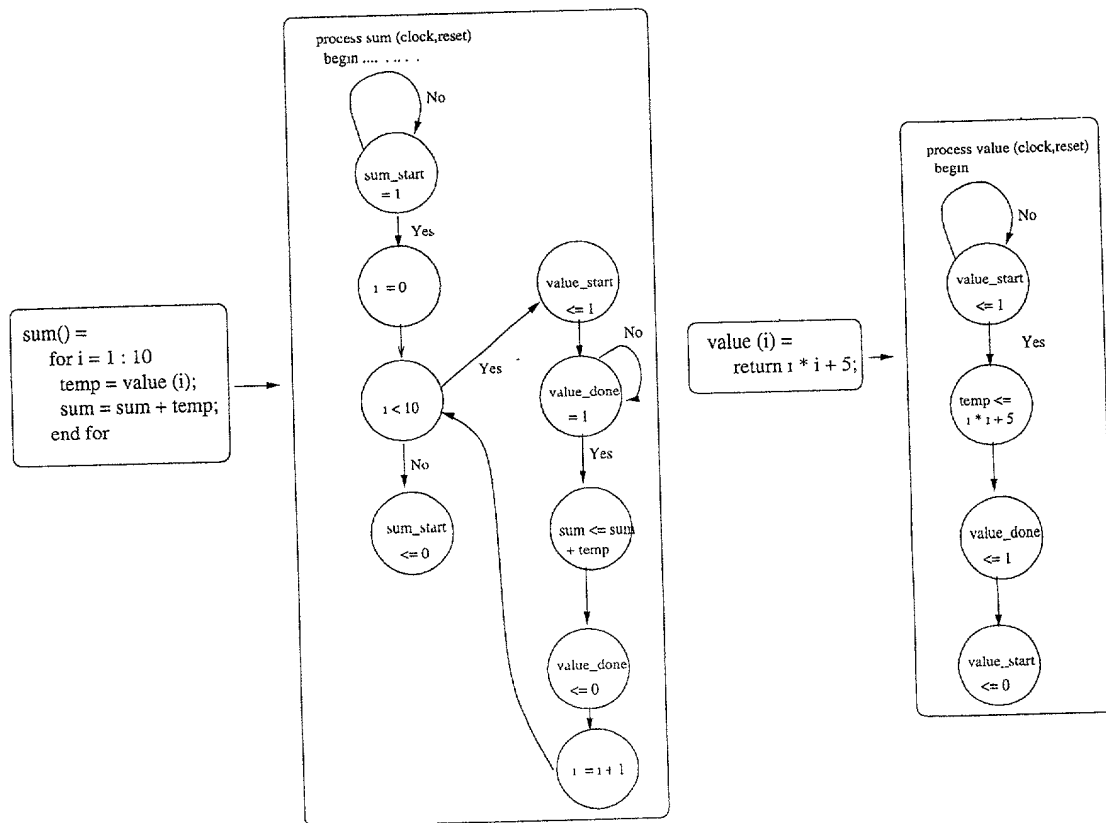


Figure 7: State machine representation of a function call in MATLAB

`a[i + 1] = b + c ;`

↓
Levelization
↓

`t1 = b + c ;`
`t2 = i + 1 ;`
`a[t2] = t1 ;`

```
when state 1 => t1 <= b + c ;
                next_state <= state 2 ;
when state 2 => t2 <= i + 1 ;
                next_state <= state 3 ;
when state 3 => mem_request <= '0' ;
                mem_write_enable <= '0' ;
                next_state <= state 4 ;
when state 4 => mem_address <= Base_a + t2 ;
                mem_data_out <= t1 ;
                next_state <= state 5 ;
when state 5 => if( mem_grant = '0' ) then
                next_state <= state 6 ;
            else
                next_state <= state 4 ;
            endif;
```

(a)

(b)

Figure 8: (a) Array statement in MATLAB and its levelization (b) VHDL corresponding to the MATLAB code. The signals *mem_request*, *mem_data_out*, *mem_grant*, *mem_write_enable* and their particular states and assignments are specified in an external file read by the compiler. *Base_a* is a constant denoting the starting address of the array *a* in memory.

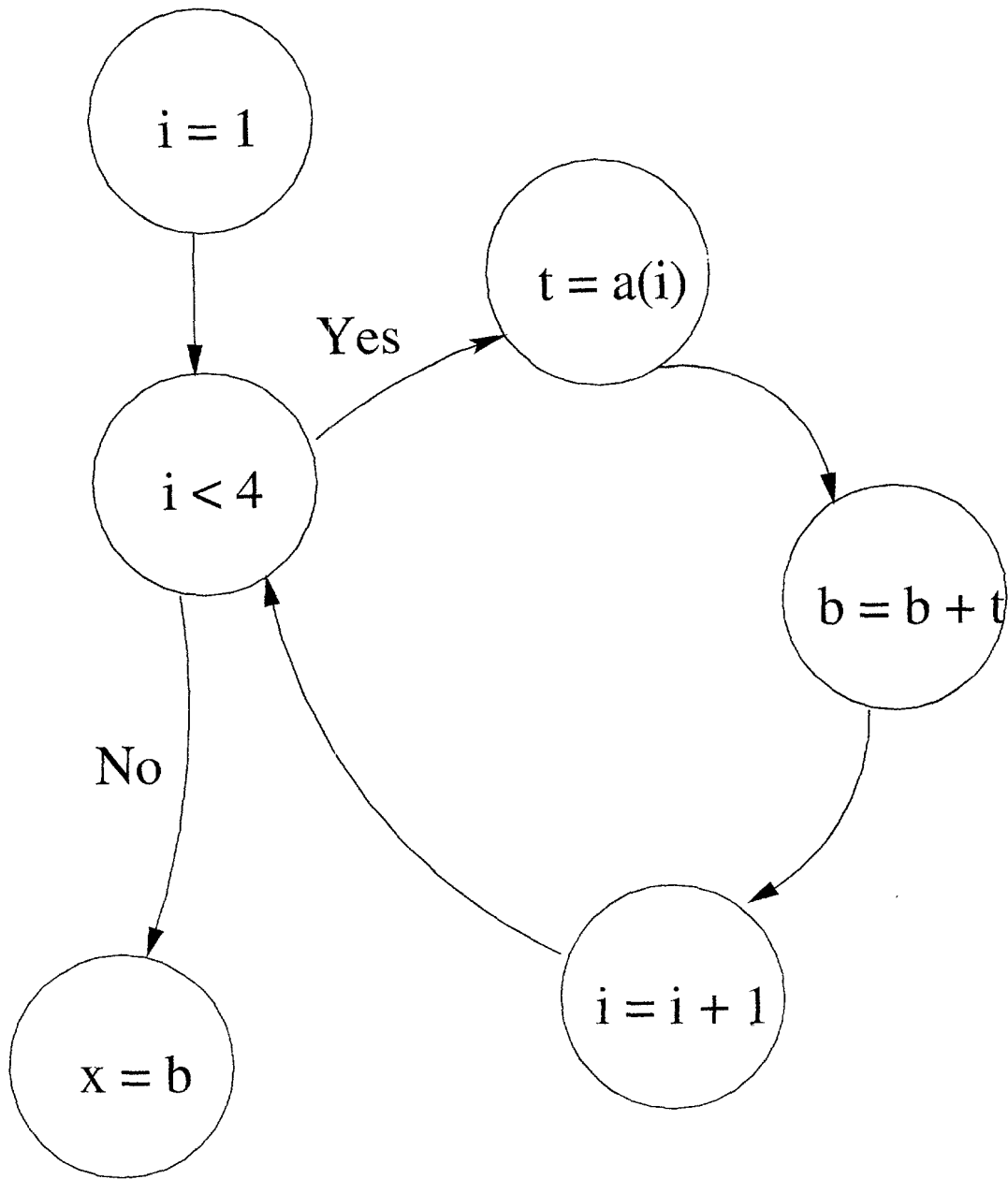


Figure 9: FSM Representation of a code section

Algorithm 1 architecture ...

```

process ...

    if reset = '1' then
        ...
    elsif rising_edge(clock)
        case control is
            when state1 =>
                i := 1;
                control <= state2;
            when state2 =>
                if (i < 4) then
                    ....
                else
                    ....
                endif;
            when state3 =>
                t := a(i);
                control <= state4;
            when state4 =>
                b := b + t;
                control <= state5;
            when state5 =>
                i := i + 1;
                control <= state2;
        end case;
    end if;
end process;

```

Figure 10: VHDL code generated for the above FSM

over manual design or design at a low level of algorithm description (e.g., directly in VHDL).

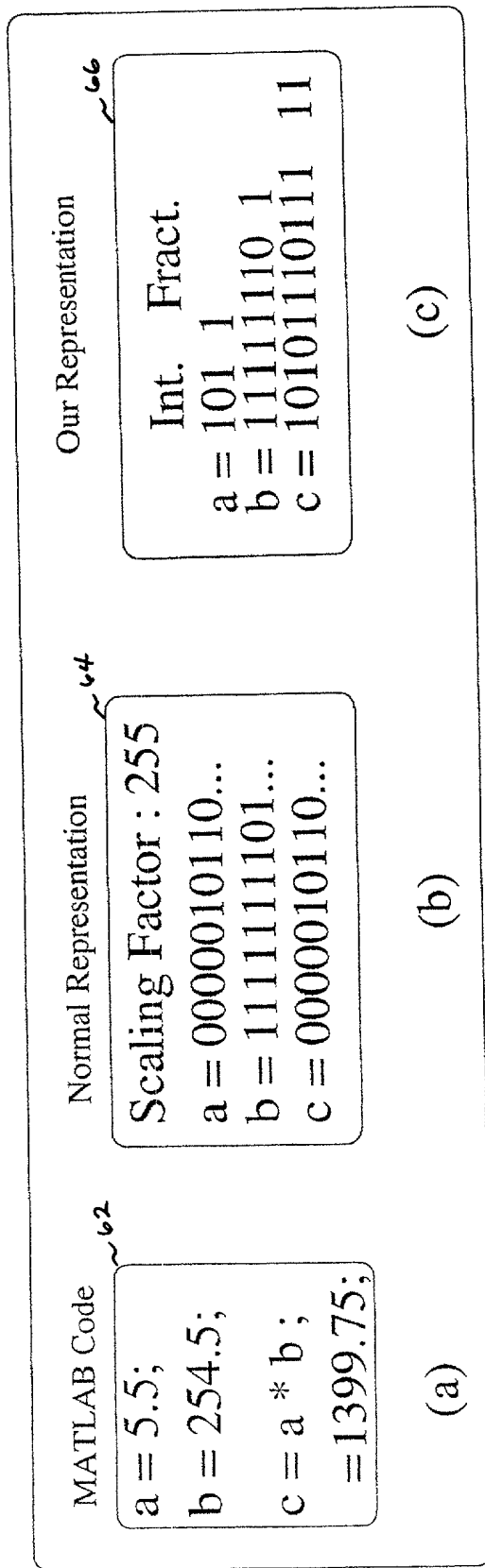


Figure 11: Representation of Real Variables

```
for i = 1:1:20  
    a[i+2] = b[i] + c[i+1];  
end
```



```
for i = 1:4:20  
    a[i+2] = b[i] + c[i+1];  
    a[i+3] = b[i+1] + c[i+2];  
    a[i+4] = b[i+2] + c[i+3];  
    a[i+5] = b[i+3] + c[i+4];  
end
```

Figure 12: Example showing loop unrolled for Memory Packing

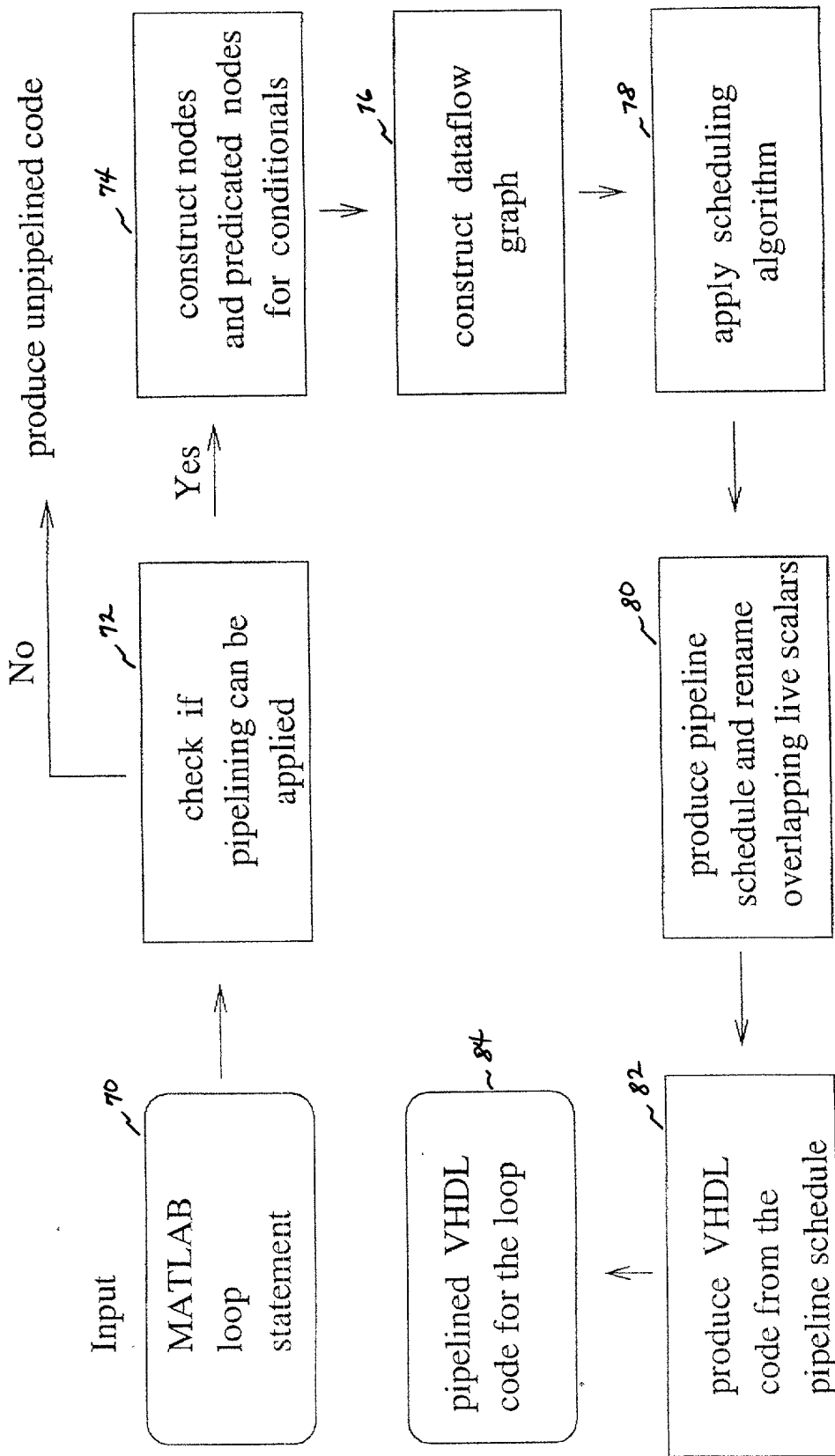


Figure 13: Overall framework for pipelining optimizations

```

for i = x : y : z
    a[i] = a[i] + 2 ;
    x    = a[i - 1] + 3 ;
end ;

```

index variable

index initial value

index step

index limit

Loop Body

Loop Statement

```

if ( i )
    a[i] = a[i] + 1 ;
    x    = a[i + 1] - 4 ;
end ;

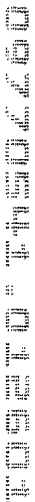
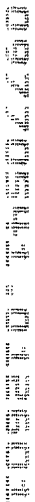
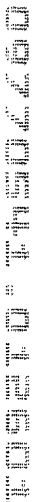
```

Conditional Variable

Conditional Statement Body

Conditional Statement

Figure 14: Illustration of Terms used in pipelining framework



Example Array Access Statement

```
x = a(i, j, k);      % a is a 256 X 256 X 256 array
```

```
s1 : ad_temp3 := i * 256 * 256 ;
```

```
s2 : ad_temp2 := ad_temp3 + j * 256 ;
```

```
s3 : ad_temp1 := ad_temp2 + k ;
```

```
s4 : array_address := BaseAddress_a + ad_temp1 ;
```

Address
Calculation

```
s5 : mem_request <= '0' ;
```

```
mem_write_enable <= '0' ;
```

```
s6 : mem_address <= array_address ;
```

Memory
Interface
Specific
Signals

```
s7
```

```
s8 : x := mem_data_in ;
```

```
s1 : mem_request <= '0' ;
```

```
mem_write_enable <= '0' ;
```

Read

```
s2 : mem_address <= [ MEMADDRESS ]
```

```
s3
```

```
s4 : [ MEMDATA ] <= mem_data_in
```

Figure 16: Example of node construction for array access statements

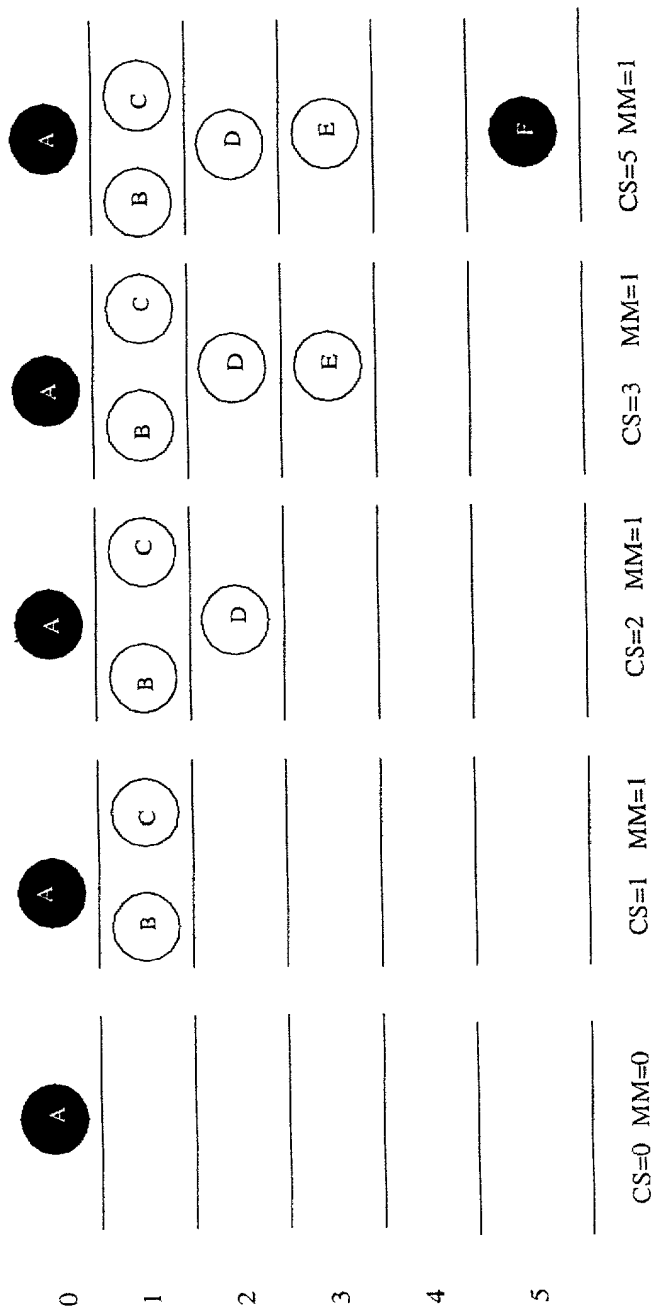


Figure 17: An illustration of pipeline method

105210-140000

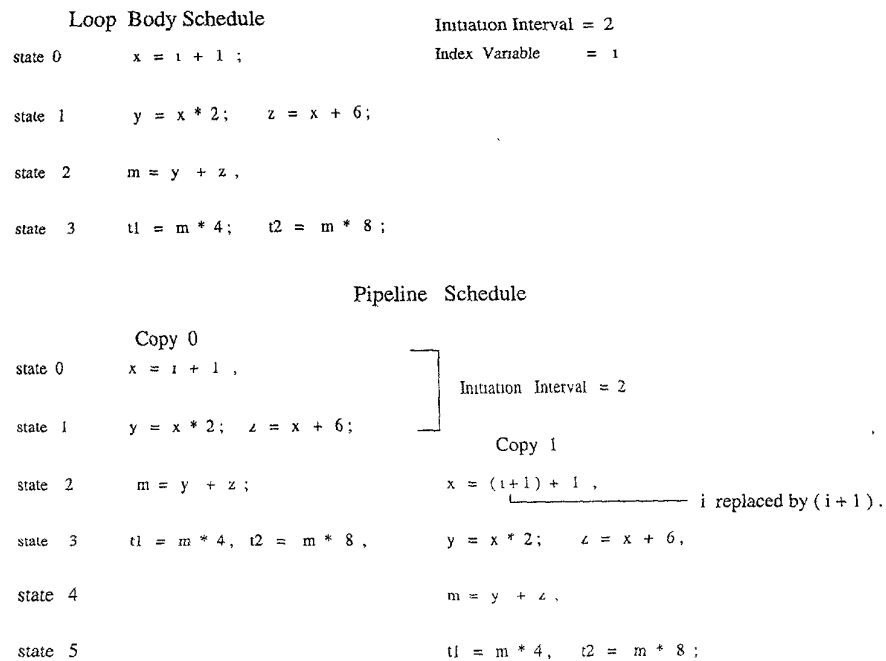


Figure 18: Construction of pipeline schedule from loop body schedule

Loop Body Schedule

```
s1  x = i + 1 ;
s2  y = i * 2 ,
s3  k = x + 3 ,
```

Pipeline Schedule

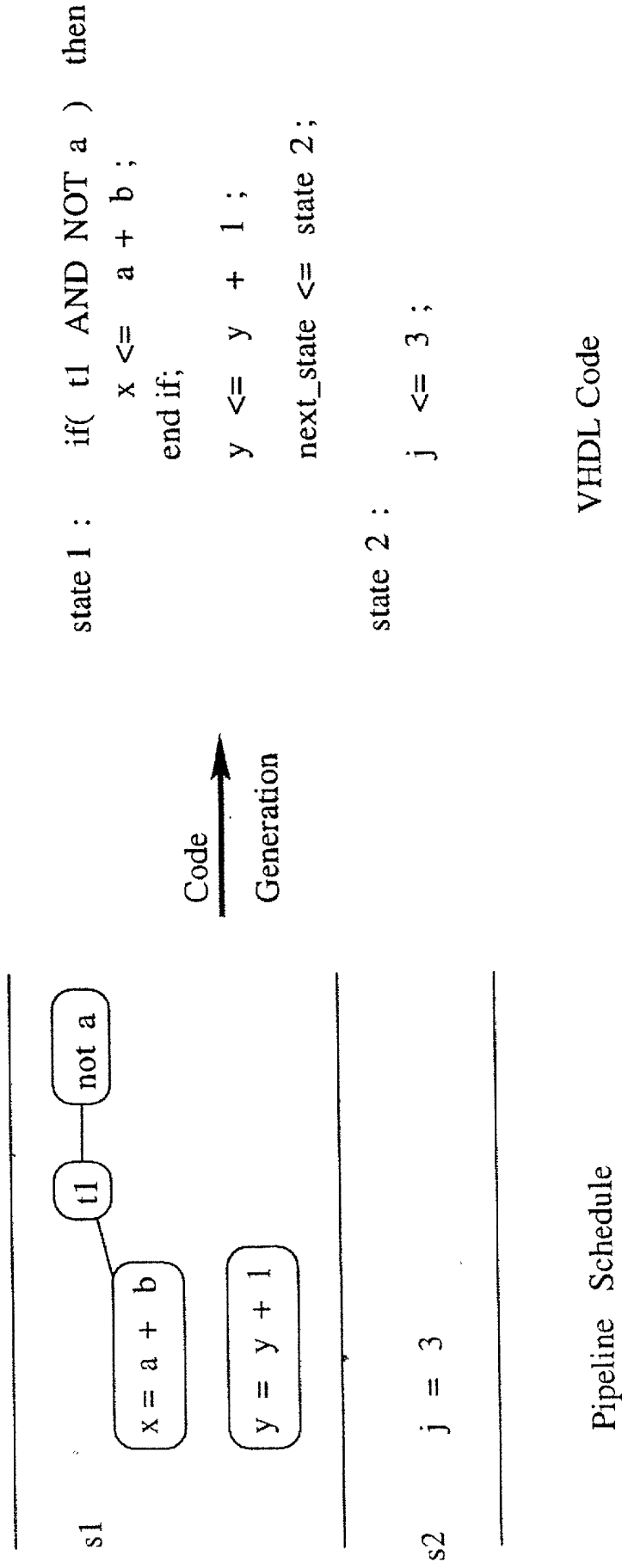
	Copy 0	Copy 1	Copy 2
s1	x = i + 1 ;		
s2	y = i * 2 ;	x = i + 1 ;	
s3	k = x + 3 ;	y = i * 2 ;	x = i + 1 ;
s4		k = x + 3 ;	y = i * 2 ;
s5			k = x + 3 ;

Vertical arrows between copies indicate 'x live range' dependencies.

Pipeline Schedule with Overlapping Live Scalars Renamed

	Copy 0	Copy 1	Copy 2
s1	case mod_var is when 0 : x0 = i + 1 ; when 1 : x1 = i + 1 ; when 2 : x2 = i + 1 ; end case ;		
s2	y = i * 2 ;	case mod_var is when 0 : x1 = i + 1 ; when 1 : x2 = i + 1 ; when 2 : x0 = i + 1 ; end case ;	
s3	case mod_var is when 0 : k = x0 + 3 ; when 1 : k = x1 + 3 ; when 2 : k = x2 + 3 ; end case ,	y = i * 2 ;	case mod_var is when 0 : x2 = i + 1 ; when 1 : x0 = i + 1 ; when 2 : x1 = i + 1 ; end case ;
s4		case mod_var is when 0 : k = x1 + 3 ; when 1 : k = x2 + 3 ; when 2 : k = x0 + 3 ; end case ,	y = i * 2 ;
s5			case mod_var is when 0 : k = x2 + 3 ; when 1 : k = x0 + 3 ; when 2 : k = x1 + 3 ; end case ,

Figure 19: Renaming of scalars with live overlapping ranges in the pipeline schedule



Pipeline Schedule

VHDL Code

Figure 20: VHDL Code generation Illustration

DENNIS S. FERNANDEZ, REG. NO. 34,160
PETER C. SU, REG. NO. 43,939

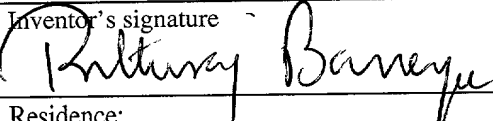
Send Correspondence and Communications to:

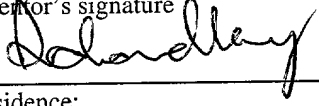
FERNANDEZ & ASSOCIATES, LLP
PATENT ATTORNEYS
PO BOX D
MENLO PARK, CA 94026-6204
(650) 325-4999
(650) 325-1203 : FAX
EMAIL: iploft@iploft.com

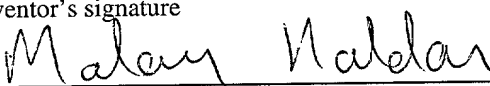


22877
PATENT TRADEMARK OFFICE

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and beliefs are believed to be true; and further that these statements were made with knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of sole or first inventor: PRITHVIRAJ BANERJEE	
Inventor's signature 	Date 1/23/01
Residence: 2130 CHANDLER LANE, GLENVIEW, IL 60025	
Citizenship: UNITED STATES OF AMERICA	
Postal Address: SAME AS RESIDENCE	

Full name of sole or first inventor: ALOK CHOUDHARY	
Inventor's signature 	Date 1/23/01
Residence: 1100 NORTH LAKE SHORE DRIVE, #74, CHICAGO, IL 60611	
Citizenship: UNITED STATES OF AMERICA	
Postal Address: SAME AS RESIDENCE	

Full name of sole or first inventor: MALAY HALDAR	
Inventor's signature 	Date 1/23/01
Residence: 2145 SHERIDIAN ROAD, L458, EVANSTON, IL 60208	
Citizenship: INDIA	
Postal Address: SAME AS RESIDENCE	

Full name of sole or first inventor: ANSHUMAN NAYAK	
Inventor's signature <i>Anshuman Nayak</i>	Date <i>1/23/01</i>
Residence: 2145 SHERIDAN ROAD, L458, EVANSTON, IL 60208	
Citizenship: INDIA	
Postal Address: SAME AS RESIDENCE	

T O S T O T O S T O T O S T O T O